



A comprehensive review of vehicle detection using computer vision

Aymen Fadhil Abbas¹, Usman Ullah Sheikh², Fahad Taha AL-Dhief³, Mohd Norzali Haji Mohd⁴

^{1,2,3}School of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

⁴Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia

Article Info

Article history:

Received Apr 27, 2020

Revised Nov 18, 2020

Accepted Dec 3, 2020

Keywords:

Computer vision

GMM

Intelligent transport system

KDE

Vehicle detection

ABSTRACT

A crucial step in designing intelligent transport systems (ITS) is vehicle detection. The challenges of vehicle detection in urban roads arise because of camera position, background variations, occlusion, multiple foreground objects as well as vehicle pose. The current study provides a synopsis of state-of-the-art vehicle detection techniques, which are categorized according to motion and appearance-based techniques starting with frame differencing and background subtraction until feature extraction, a more complicated model in comparison. The advantages and disadvantages among the techniques are also highlighted with a conclusion as to the most accurate one for vehicle detection.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Usman Ullah Sheikh

School of Electrical Engineering

Faculty of Engineering

Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Email: usman@fke.utm.my

1. INTRODUCTION

Recently, video cameras have been extensively used for traffic surveillance applications, which effectively provide valuable information about traffic flow. Additionally, the usage of video cameras for traffic surveillance is complemented by the incorporation of advanced technology, for example, computer graphics, state-of-the-art cameras, high-end computing capabilities, and automatic video analysis [1]. For road safety and security, detection of vehicle of make and model has become necessary for the identification of vehicles on urban roads. Vehicle detection is the key for numerous applications related to the surveillance camera but vehicle detection comprises multiple proven techniques which this paper, describes in detail [2, 3].

For surveillance systems, one of the most popular types of devices are visible spectrum cameras, which have been extensively used for monitoring environments, events, activities, and people. Different studies in background-foreground segmentation [4–7] and object detection classification [8–10] have also attempted to automatically analyze image or video data from surveillance cameras.

Another method that is fast gaining popularity because of the significantly increased road transport levels especially on urban roads, is the intelligent transportation system (ITS). The aim of ITS are to improve driving safety and security, thereby making roads more flexible even during traffic congestion [2]. This is achieved by applying the principles of computers, electronics and communication to various systems, and involving vehicles, drivers, passengers and managers in all inter-communicate framework. ITS utilizes visual

appearance to detect, recognize, and track vehicles which may facilitate the detection of incidents and analyze behavior [1, 11]. It is also able to retrieve data on traffic flow parameters, such as vehicle class, count and trajectory.

A crucial part of ITS involves vehicle detection, with much work being done on the topic, but there are still challenges facing the effective detection of vehicles from the front or rear, illumination and shadows affecting the detection of vehicles and occlusion issues. Many recent studies have attempted to develop appearance-based techniques to detect vehicles just from still images. This paper seeks to provide a comprehensive review of the various techniques, particularly those in vehicle detection, that could be available for video-based traffic surveillance from a computer vision perspective.

2. VEHICLE DETECTION

The first step in video-based analysis for ITS is vehicle detection [1], which is very important for vehicle recognition, tracking and enhanced processing [12]. Figure 1 shows the main steps of the ITS system. Research into the field is further subdivided into two categories: motion-based and appearance-based techniques [13]. Motion-based techniques utilize motion cues to differentiate among moving vehicles from stationary background. Appearance-based techniques utilize appearance features of vehicles such as color, shape and texture to isolate the vehicle from the surrounding background scene.

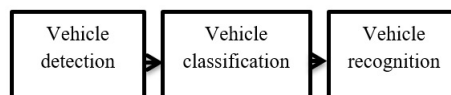


Figure 1. General diagram showing the main steps of intelligent transportation system

2.1. Motion segmentation

Motion segmentation, a crucial step in computer vision algorithms, seeks to decompose video into a series of moving objects and background. For computer vision algorithms, the first important step is decomposition. Various applications rely on this, including metrology, video surveillance, inspection and robotics among others. Various studies have highlighted the problem of segmentation, but the performance is not perceptible enough for human beings [14, 15]. Motion segmentation could be categorized either as background subtraction or optical flow [16].

2.1.1. Background subtraction

Background subtraction, studied since the 1990s, is employed to detect moving objects in videos without prior knowledge. It is mainly employed for video-surveillance applications, as the need for detecting people, animals and vehicles arise before more complicated processes such as intrusion detection, tracking and people counting could be undertaken [17, 18]. Background subtraction techniques could be categorized as parametric, non-parametric or predictive.

2.1.2. Non-parametric background modeling

These techniques employ the pixel history of images to develop probabilistic representations of perceptions [19]. Examples of this technique are Kernel density estimation (KDE), codebook model and background extractor (vibe). Background modelling, while still difficult to use in complicated scenarios, is crucial to video surveillance. Difficulties arise from illumination variance and dynamic backgrounds.

2.1.3. Kernel density estimation (KDE)

Kernel density estimation (KDE) is a nonparametric technique which attempts to approximate density, whereby a known density function (i.e. the kernel) is taken as the mean within perceived data points creating fluid estimations [20]. A novel technique for background modeling based on adaptive non-parametric kernel density estimation (AKDE) proposed [21]. In this technique, base-line system is identified to solve the problem of detecting foreground regions in videos with semi-stationary background. AKDE deals with the multi-modality of the background as well as scene independence. The benefits of AKDE are due to global

threshold not being required for pixels in video scene, different and adaptive thresholds for individual pixels. Instead of altering parameters, the system is able to reliably work with various video scenes.

Two innovative techniques are also proposed for background modeling based on non-parametric density estimation and recursive modeling. For novel recursive modeling method (RM) updates, the model is updated online when a new frame is present rather than processing video frames to generate background models. The method is very robust (accuracy of roughly 84%) compared to other techniques, due to its rapid and non-periodic nature. The caveat to the method is that it may fail if no probabilistic model for background or foreground is present.

Another researcher, [22], proposed an efficient background subtraction framework dealing with variance in illumination on the feature level, which also employed local binary pattern (LBP) histograms for background modeling. Furthermore, LBP was extended to a scale invariant local ternary pattern (SILTP) operator. Precise segmentation of dynamic objects could be accomplished by modeling pixel process with single local patterns. The problem is that local patterns are not ordinal numerical values and cannot be modeled immediately to traditional density functions. This leads to the proposal of a pattern kernel density estimation (PKDE) to effectively model probability distributions. Complex scenes data sets are proposed and contrasts with current online background subtraction algorithms like Gaussian ("MoG"), pixel-wise LBP histogram based (LBP-P) and blockwise LBP histogram-based approach (LBP-B), having achieved over 92.40. The prime disadvantage with the method, however, is the inability to work with color video.

A KDE technique over a joint domain-range representation of image pixels proposed [23]. This technique could model spatial relations among observed intensities using a joint domain-range representation of pixels. Using samples from previous frames, background and foreground processes may be modeled with non-parametric density estimation. Joint domain-range KDE shows the background and foreground processes by bringing together three color and two spatial dimensions within a five-dimensional joint space. Using this method, accuracy of 71.66% was reached, but bore the disadvantage of the foreground priors not being able to be modeled for complicated objects and also does not possess object tracking information.

KDE has also previously been employed to track vehicles with video from vertically placed camera [24]. Foreground objects are tracked with automatic adaptation to different environments. Evolutionary computing framework is also employed for pose estimation by recovering the pose of vehicles in subsequent frames, with the deformable modeling being fitted into the image data to process the occlusion. Using this method, the accuracy was found to be 88.5%, with the limitation arising due to the vehicle not being able to be tracked without the shadow removal algorithm and inability to be processed on RGB frames directly. KDE also has difficulties in detecting too many objects in the same scene.

2.1.4. Codebook model

The Codebook model considers a number of dynamically handled codewords to model background pixels, by substituting parameters of probabilistic functions using pixel textures, pixel color and region experience [25]. Foreground region is detected using clustering of texture information. Color cues from background are also modeled by the codebook and utilized for refinement of texture-based detection results derived from color and texture traits. Using this method, the accuracy was found to be 69%, with the limitation being the detection of foreground regions when both texture and color are the same as background scene. In another study, a hierarchical background subtraction algorithm consisting of block-based stage and pixel-based stage was proposed [26]. In block-based stage, apparent backgrounds are detected with block-based codebook, while maintaining spatial relations among questionable foreground pixels. This way, the accuracy of this work was found to be only 60%, which is low compared to that obtained by other studies. There was also a limitation in being able to detect foreground in RGB space.

Another method to detect the moving vehicles in crowded traffic scenes was a Gaussian pyramid layered algorithm and codebook in conjunction with Local binary patterns to construct low-resolution and high-resolution for the Gaussian pyramid [27]. Using this method, the accuracy was found to be 69.861%, which is comparatively high. In summary, the codebook technique is more advanced than KDE, especially when dealing with foreground scene, as it can readily distinguish between background and foreground details. The main limitation is in the low accuracy when a scene has a similar color in RGB space to background.

2.1.5. Visual background extractor (ViBe)

The visual background extractor (ViBe) is a robust method employed to extract background with relatively high accuracy and lower computational cost. In this method, the novel usage is with random policy

to choose values upon which to build sample-based background estimation [28]. A study proposed a method which considers a universal sample-based background subtraction algorithm, used to store a set of values for each pixel taken beforehand [28, 29]. A comparison is then made to the present pixel value to detect whether the pixel belongs in the background and adjusts the model accordingly. It then compares this set to the current pixel value in order to determine whether that pixel belongs to the background, and adapts the model. With random selection of values to substitute from background, this method is novel from classical methods which consider the oldest values being replaced first. Table 1 shows all the techniques that are used in non-parametric model along with their accuracy and limitations.

Table 1. Non-parametric background model

Method	Accuracy	Gap
KDE [21]	84%	Fails in performing the foreground/ background segmentation without establishing a probabilistic model for the background or the foreground.
KDE [22]	92.40%	Fails to work on color information.
KDE [23]	71.66%	The foreground priors cannot be modeled if there is complex object in the scene and cannot include the object tracking information.
KDE [24]	88.5%	Vehicle cannot be detected without shadow removal algorithm and cannot work on RGB frames directly.
Codebook model [25]	69%	Fails to detect foreground regions when their properties of both the color and the texture are similar to those of the background scene.
Codebook model [26]	60%	Fails to achieve high accuracy on RGB color space.
Codebook model [27]	69.861%	Did not consider the foreground scene or multi objectives on the foreground.

2.1.6. Parametric background modeling

Parametric methods consider the pixel intensity (i.e. pixel color) being modeled by employing a probability distribution with a known parametric form [30]. A popular method for defining the mean and variance as parameters is the Gaussian distribution, which estimates parameters with temporal averaging or temporal nonlinear filtering when a single Gaussian distribution is used for background models [31]. However, these methods are not applicable to complex dynamic scenes whereby background pixels could vary greatly in color distribution. Table 2 shows the works based on parametric background model.

2.1.7. Frame averaging

In traditional averaging techniques, a set of N frames are summed up and divided by the number of frames [32]. The resulting model is then subtracted from subsequent frames, a popular method used by other studies [32, 33]. The accuracy was shown to be 95% in one study [32] but could not perform with effects of shadows, while another study showed accuracy of 97% while still suffering from the same problem [33].

2.1.8. Single Gaussian

Temporal single Gaussian is used to model background iteratively, enhancing performance and reducing memory used. A single Gaussian background model is used for foreground segmentation, with every segmented foreground silhouette, followed by a feature vector based on silhouette measurements as calculated [34]. Foreground pixels which are part of a dynamic object are readily detected using an adaptive background subtraction scheme, whereby each background pixel is modeled as an individual Gaussian process with mean μ (time-averaged intensity) and σ (standard deviation of intensity) [35].

2.1.9. Sigma-delta

First shown by A. Manzanerra [36], the sigma-delta background estimation seeks to compare between elementary increments and decrements. Using this method, accuracies as high as 95% have been achieved, with false alarms of only 45%. Another study by S. Toral showed improvements of stability [37], with the enhanced method selectively recalculating background pixels to successfully detect slow moving vehicles. Another study considered a novel algorithm for queue-parameter estimation by utilizing two background models estimating temporal stops in vehicles [38]. The method was found to be computationally efficient, with accuracy of 98%, while the main limitation was in detecting a vehicle when another vehicle passes by the same point.

2.1.10. Gaussian mixture model

Gaussian mixture model (GMM) models two or more Gaussians temporally with online updates [39]. In one study, GMM was utilized for background subtraction [40], but yielded noisy images due to false classification. The method was enhanced with hole filling algorithm (HF), with resulting accuracy of 97.9%. Other studies have achieved accuracy of 91.26% [41] and 97.22% [42] 79.63% in heavy traffic conditions), using the same method. Another study [43] found the extraction speed of this method not capable enough with resulting noisy blocks. In an attempt to solve the issue, a vehicle ROI extraction method based on area estimation GMM was developed, with results showing detection speed per frame to be 30 (frames/sec), having tested on 537 vehicles. This result is quite convincing in terms of speed and ability to filter out noise.

Table 2. Parametric background model

Method	Accuracy	Gap
Frame averaging [32]	95%	Fails under effects of shadows.
Frame averaging [33]	97.07%	Requires development of extraction of additional features such as vehicle class, speed and inter-vehicle distance to reduce system delay
Single Gaussian [34]	Bike 71.7% , car 93.2%, van 100.0%, bus 93.8%	Requires incorporation of local feature information into the classifier to have a second source of information apart from the motion foreground.
Single Gaussian [35]	98% south lane , 93% north lane	Low accuracy during extraction of foreground features.
Sigma-delta [36]	Accuracy 95%,false alarm 45%	Failure to detect foreground.
Sigma-delta [38]	98% full queue	Failure to classify two cars as the same object because one car entered the scene simultaneously while another car left at the same point.
GMM [39]	Process 11 to 13 frames a second (frame size 160x120pixels, Every object which entered this scene – in total, 33 cars and 34 people – was tracked. It successfully classified every car	Requirements to adjust the value of the radius k manually by guessing the distance between the neighborhood pixels to vote every pixel in the scene which might get error and effect badly on the accuracy.
GMM [40]	97.92%	Requirements for more processing time in highly traffic conditions.
GMM [41]	91.26%	Failure to detect vehicles in heavy traffic scene.
GMM [42]	light traffic 97.22% heavy traffic 79.63	Lower accuracy during heavy traffic scene.
GMM [43]	Actual number of vehicles is 541, number of vehicle that the ROI includes 537 ,Number of false detection vehicle ROI 65 ,Detection speed Frames/S30	Effective filter out ambient noise but requirements to increase processing speed further and use more vehicles during testing.

2.1.11. Predictive background modeling

Predictive procedures are employed in modeling the state dynamic of each pixel. Examples include Kalman filtering [44, 45] and eigenspace reconstruction or eigenbackground [46].

2.1.12. Kalman filter background modeling

The Kalman filter, developed by Kalman and von Brandt, is a recursive estimator which computes current measurement and estimated state in order to approximate the present state [47]. While this method could approximate background images (with each filter representing one color), foreground could potentially be noisy for filter states, and changes in illumination are non-Gaussian noise which disrupt the utilization of Kalman filters. The method is still effective for dealing with multiple objects for tracking rather than detecting as the latter requires separating background scene from foreground.

One study by [48] developed a technique for detecting, tracking and counting vehicles using Kalman filtering, where the background was updated using the method to detect, track and count vehicles. Accuracy of the method was found to be 97.50%, with the main disadvantage being the difficulty in managing moving background pixels. Another study [49] proposed a method for detecting dynamic objects using Kalman filtering, with results showing values of MSE ≤ 10 and values of PSNR. Furthermore, correlation coefficient was

found to be higher than 75%. The limitation of the study was inability to recognize very small objects, thereby requiring the initiation of multimodal features to identify objects.

2.1.13. Eigenbackground

Eigenbackground is a robust method for background subtraction, especially in crowded scenes with slow moving foreground objects. The method is also relatively computationally efficient. One study by Oliver *et al.* investigated background modeling for offline learning and online classification [46]. The mean of sample frames set was computed and subtracted from all frames, after which the covariance matrix was computed and eigenvector matrix composed from the optimal eigenvectors. For classification, the frames are projected onto the eigenspace and back projected onto image space to determine background models.

Using this method, the study reported accuracy of 68.7% for hidden Markov models (HMMs) and 90.9% for coupled hidden Markov models (CHMMs). Another study considered block-level eigenbackgrounds, whereby the original frame was divided into blocks to perform independent block training and subtraction [50]. In [51] improved the algorithm further for choosing optimal eigenbackgrounds for individual pixels by selective training, model initialization and pixel-level reconstruction. In [52] proposed another variant of this method targeted towards crowded scenes. Using selective mechanisms for background modeling and subtraction, virtual frames were generated as training and updated samples for the eigen backgrounds. This way, optical background models could be initialized, called selective model initialization and the best Eigen background could be chosen.

Thus far, all methods that have been discussed in background subtraction section show that background modeling is a widely used approach for detecting moving objects from static cameras. It is the general way of motion detection. It is a process that finds the difference of the current image and the background image to detect the motion region, and it is commonly efficient to deliver data included object information. In Table 3, all the recent predictive techniques are illustrated. To be clearer about this technique, the latest techniques employed by various study results are shown to show their methodology. The limitation of the background subtraction technique is still its low accuracy in crowded scenes and also the tendency to often fail if the foreground scene possesses multiple objects. But the main disadvantage of background subtraction and the most questionable one is that it works only on videos frames and cannot work on still images. This is why recently, especially in last ten years, the direction of vehicle detection employed feature-based techniques because due to its ability to work on still images.

2.1.14. Optical flow

Optical flow detects motion with analysis of pixel motion, whether as a group or individually. The concept of optical flow is derived from how human beings observe motion of objects by capturing and analyzing gradient, brightness, lights reflected and so on [53]. The method was proposed by [54] who employed flow vector characteristics for dynamic objects over a period of time to detect dynamic regions in video. The instantaneous pixel speed on image surface relates to a particular object in the 3D space. Results from [54] were displayed in terms of the average hit rate, with results of 71.83%, 87.21% and 87.04%, respectively for window sizes of 5 by 5, 10 by 10 and 15 by 15, respectively. The main limitation was found to be that some objects were incorrectly detected as vehicles.

Another study considered the parallel optical flow based on Lucas-Kanade algorithm [55]. In this method, the foreground optical flow detector detected the object with binary computation conducted to define rectangular regions surrounding the detected object. The findings showed success rate of 98%, which was determined accurate for front facing camera. The study reported limitations for vehicles being missed by optical flow detector in heavy traffic flow conditions, where frames contained too many images in the same scene.

In [56] proposed moving vehicle detection algorithm based on optical flow estimation on the edge image. What was unique about the method was the implementation of the pyramid model of the Lucas-Kanade optical flow to calculate optical flow information for a feature point set. However, the method still persisted from the same limitation of inability to detect vehicles effectively in heavy traffic flow conditions for the same aforementioned reason. The summary of predictive background models is shown in Table 3.

Table 3. Predictive background model

Method	Accuracy	Gap
Kalman filter [45]	Track a car in a campus. The car come into the scene, then stayed at a carport. The resolution of the test video is 384*288=	Small data set.
Kalman filter [48]	97.50%	Difficulties in managing dynamic background pixels
Kalman filter [49]	This method min value of MSE is 2.90 and maximum value of MSE is 8.40, which are less than 10. Min value of PSNR is 84.83 and max value is 95.30, which are more than 75%. Min value of Correlation coefficient is 90.6 and max value is 95.0, which are more than 75%. Min value of Similarity is 90.11 and max value is 95.44, which are more than 75%.	Inability to recognize the very small objects, thus requiring multimodal features to characterize objects.
Eigenbackground [46]	HMMs =68.7 , CHHMs = 90.9	Low accuracy.
Eigenbackground [52]	The object detection result shows the output image we can observe the object without background	Low accuracy and requirement of further objects to interact in one scene.
Optical flow [53]	98%	Vehicles were missed by the optical flow detector due to heavy traffic flow conditions.
Optical flow [56]	The vehicle velocity with high accuracy of about 0.212 cm/sec and within a Percentage error of 0.986%.	Fails to detect front view vehicle.

2.2. Appearance based techniques

Whereas motion segmentation methods could detect only motion, appearance-based methods seek to detect stationary objects in either image or video by way of their features. Utilization of visual information, such as color, texture and shape for vehicle detection still needs pre-requisite information [57]. Thus, feature extraction is used to compare between 2-D images with real-world 3-D images.

2.2.1. Feature based techniques

Feature-based techniques employ coded descriptions for the features, in order to describe the visual appearance of vehicles. Various features have been used in vehicle detection, including local symmetry edge operators [58]. This method is sensitive to size and illumination variations, thus requiring more spatial invariance edge-based histograms, as was formerly applied by [59]. Recently, features have evolved into more advanced ones enabling direction and classification of vehicles.

2.2.2. Haar-like features

Haar-like features refer to the formation of sum and difference of rectangles for image patch, describing grey-level distribution of adjacent regions [60] Filters used to extract features comprise two, three or four which could be positioned and scaled accordingly. Output of filters are computed by adding the pixel values for the grey region and white region one by one, and then normalizing the difference between two.

These features have been formerly used to detect vehicles from still cameras. Furthermore, statistical data could be computed, including average traffic speed and flow. The framework comprises three main stages, including first detecting the vehicles using Haar-like features. Several studies have employed training of cascaded Adaboost classifier [61, 62], while others detected highway vehicles [63] and urban vehicles [65]. One study employed LBP features to train boosting classifiers for vehicle detection for license plates [?], while another study used MLP ensemble in conjunction with Haar-like features [66].

In general, Haar-like features were found to possess high computational efficiency, while still being sensitive to vertical, horizontal and symmetric structures which increases their applicability for practical applications. Table 4 details the most relevant and recent studies on the topic. One study employed the method for a bicycle, concluding that it could still not perform for a greater number of vehicles [62]. Another study [65] found the accuracy for vehicle detection to be 80% with 120 false alarms resulted.

Table 4. Feature based techniques

Method	Accuracy	Gap
SIFT [57]	matching locations and scales (Match %) =78.6 match in orientation (Ori%)=71.8	Need to work on building models from multiple views that represent the 3D structure of objects.
SIFT [69]	94.13%	Fails in processing speed due to slow time processing.
SIFT [68]	87%	Low accuracy.
SURF [?]]	Average accuracy using FID scheme :94.94%	Fails in occlusion, bad lighting and non-front view.
SURF [70]	A to B camera :72.92% B to C camera :80.32%	Low accuracy and slow.
HOG [71]	Classification accuracy = 92.1%	Requires work on classifier with frame to frame tracking that will provide the opportunity to demonstrate the full potential of the algorithm on partially occluded objects.
HOG [72]	Test accuracy = 93%.	Errors in the system due to inaccurate hypotheses generation.
HOG [73]	average vehicle detection rate of 97%	The calculation costs could potentially be reduced to a range of real-time processing for autonomous driving.
HOG [74]	detection result in traffic video 1 (RDHOG) accuracy = 95.26 detection result in traffic video 1 (RDHOG) accuracy = 99.72.	Fails to detect vehicles in complicated scenes with multiple occluded objects.
Haar-like-features [61]	Bicycle: SAOT (95%), CT (64%), RDET= (64%), IVT= (64%).	Fails in occlusion and large number of vehicles.
Haar-like-features [65]	80% accurate detection with 120 false alarm.	Small number of vehicle used for detection.

2.2.3. SIFT (Shift-invariant feature transform)

David G. Lowe proposed the shift-invariant feature transform (SIFT) to transform images into a huge number of local feature vectors [57]. Each local feature was the same for image translation, scaling and rotation, with minor differences in illumination and 3D projection. The main limitation given in the study was the need for multiple views of the 3D structure for representation, Some works have been done using simple SIFT feature in [75, 60?]

In another study, X.Chen [69] proposed a method for UAVs using the SIFT method in conjunction with the implicit shape model (ISM), finding a detection accuracy of 94.14% from a training sample size of 1,110. The main limitation was its slow processing speed. [68] utilized a similar SIFT method resulting in a detection accuracy of 87% which is still considered relatively low. Thus, from the aforementioned studies utilizing SIFT, the conclusion was made in the studies that SIFT is computationally expensive, as the gradients of individual pixels need to be computed.

2.2.4. SURF (Speeded up robust features)

The speeded up robust features (SURF) is a scale and rotation invariant interest point detector and descriptor which, compared to SIFT's computational complexity, is much lower due to the substitution of Gaussian filter with a box of filters which slightly affect performance [78]. This algorithm employs a Hessian matrix approximation on integral image to local the points of interest, with second-order partial derivatives describing local curvatures [79].

Works in [?]] and [?]] both employed this method for vehicle detection, finding vehicles per frame to be 21 (frames/sec). In another study [70], a GPU based multiple camera was used to match unique representations of vehicles. The accuracy was found to be 94%, with limitations including rotation instability, illumination variations and occlusion. The low accuracy and computational time were suggested to be improved for future studies.

2.2.5. HOG (The grid of histogram of oriented gradients)

The grid of histogram of oriented gradient (HOG) [80] was initially meant to detect pedestrians but was later expanded for vehicles using 3-D model surface instead of 2-D grid of cell to generate 3-D histogram

of oriented gradient [71]. It calculates image gradient directional histogram, an integrated presentation of information pertaining to gradient and edge. Studies have shown accuracies of 92.1% [71], 93% [72] 94% [73] and 95.26% [74], with the main limitation given that the method needs to track the classifiers frame by frame, particularly for partially occluded objects [81]. Nevertheless, the improved computational efficiency compared to SIFT is still noted in the studies [82, 83].

3. DISCUSSION

Most of the studies done have dealt with highway and urban roads and the most popular techniques used in recent years were either background subtraction or feature-based techniques, and especially SIFT, SURF and HOG. Despite all these techniques, there are some issues that have negatively impacted such systems. Examples of such challenges include camera view and operating condition, which present additional limitations. ITSs face a number of difficulties, especially with urban roads traffic scenes and intersections in which heavy traffic, vehicle occlusion, and camera placement affect the performance of the systems. These challenging issues still require additional research and development, especially in the case of urban roads. In recent years, studies have focused on how to detect vehicles in complex scenes and how to detect vehicles under occlusion objects or in poor lighting situations. Researchers and also tried to use huge vehicles datasets in their efforts to achieve high accuracy.

4. CONCLUSION

This paper has presented different vehicle detection techniques in video-based traffic surveillance and monitoring systems. Vehicle detection falls into two major branches: motion-based and appearance-based techniques. Both techniques can be applied to isolate vehicles from the background scene with varying computational complexity and detection accuracy. Following the descriptions of various existing techniques, it can be observed that most methods achieved high levels of accuracy. However, most of detection systems still have some limitations which negatively impact the accuracy levels. These limitations include poor lighting, weather changes, occlusion, and shadows. As a future direction for further research on this topic of vehicle detection, most researchers will focus on detecting vehicles under occlusion and detection of the front and rear view of the vehicle under poor lighting conditions while removing the shadow from the scene.

ACKNOWLEDGEMENT

The authors would like to thank the Research Management Centre, Universiti Teknologi Tun Hussein Onn (UTHM) for its support under the Research Fund E15501, the Ministry of Education Malaysia and Universiti Teknologi Malaysia (UTM) for their support under the Research University Grant (GUP), grant number 19H61.

REFERENCES

- [1] M. Swathy, et al., "Survey on vehicle detection and tracking techniques in video surveillance," *International Journal of Computer Applications*, vol. 160, no. 7, pp. 22-25, 2017.
- [2] J. J. Antony, M. Suchetha, "Vision based vehicle detection: A literature review," *International Journal of Applied Engineering Research*, vol. 11, no. 5, pp. 3128–3133, 2016.
- [3] R. D. Sharma, S. K. Gupta, "Moving object detection and tracking based on background subtraction," *Object Detection, Classification, and Tracking Technologies*, vol. 2018, Oxford Journal of Intelligent Decision and Data Science, 2018, pp. 55–62.
- [4] S. Jeeva, M. Sivabalakrishna, "Survey on background modeling and foreground detection for real time video surveillance," *Procedia Computer Science*, vol. 50, pp. 566–571, 2015.
- [5] N. C. Joy, J. Prasad, "Survey on background modeling and foreground detection methods," *International Journal for Innovative Research in Science Technology*, vol. 3, no. 4, pp. 338–41, 2016.
- [6] M. Shehata, et al., "Vehicles detection based on background modeling," *International Journal of Engineering Trends and Technology*, vol. 66, no. 2, pp. 92-95, 2018. Available: arXiv:1901.04077.
- [7] S. Joudaki, et al., "Background subtraction methods in video streams: a review," *2015 4th International Conference on Interactive Digital Media (ICIDM)*, IEEE, 2015, pp. 1–6.
- [8] P. M. Tank, H. A. Patel, "Survey on human detection techniques in real time video," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 7, no. 4, pp. 5852–5858, 2018.

- [9] P. S. Sivaraman, M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE transactions on intelligent transportation systems*, vol. 14, no. 4, pp. 1773–1795, 2013.
- [10] C. V. Priscilla, S. A. Sheila, "Pedestrian detection-a survey," *International Conference on Information, Communication and Computing Technology*, Springer, 2019, pp. 349–358.
- [11] G. Sreenu, M. S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *Journal of Big Data*, vol. 6, no. 1, pp. 48–74, 2019.
- [12] N. Buch, S. A. Velastin, J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011.
- [13] B. Tian, B. T. Morris, et al., "Hierarchical and Networked Vehicle Surveillance in ITS: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 25–48, 2017.
- [14] M. Anton, J. Anitha, "A survey of moving object segmentation methods," *International Journal of Advanced Research in Electronics and Communication Engineering*, vol. 1, no. 4, pp. 73–80, 2012.
- [15] Q.-L. Li, J.-F. He, "Vehicles detection based on three-frame-difference method and cross-entropy threshold method," *Computer Engineering*, vol. 37, no. 4, pp. 172–174, 2011.
- [16] R. Manikandan, R. Ramakrishnan, "Video object extraction by using background subtraction techniques for sports applications," *CiiT International Journal of Digital Image Processing*, vol. 5, no. 9, pp. 435–440, 2013.
- [17] S. Brutzer, et al., "Evaluation of background subtraction techniques for video surveillance," *CVPR 2011*, IEEE, 2011, pp. 1937–1944.
- [18] A. Sobral, A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.
- [19] C. Cuevas, N. Garcia, "Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies," *Image and Vision Computing*, vol. 31, no. 9, pp. 616–630, 2013.
- [20] Y. Sheikh, M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [21] A. Tavakkoli, et al., "Non-parametric statistical background modeling for efficient foreground region detection," *Machine Vision and Applications*, vol. 20, no. 6, pp. 395–409, 2009.
- [22] S. Liao, et al., "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2010, pp. 1301–1306.
- [23] M. Narayana, et al., "Improvements in Joint Domain-Range Modeling for Background Subtraction," *BMVC*, vol. 1, Citeseer, 2012, pp. 5–15.
- [24] R. Khilar, S. Chitrakala, "A Novel Method for Effective Vehicle Detection and Tracking Eliminating Occlusion," *Asian Journal of Research in Social Sciences and Humanities*, vol. 6, no. cs1, pp. 714–727, 2016.
- [25] S. Noh, M. Jeon, "A New Framework for Background Subtraction Using Multiple Cues," *Asian Conference on Computer Vision*, Springer, 2012, pp. 493–506.
- [26] S. S. Shri, L. Latha, "Non-Parametric Background Modeling for Foreground Extraction using Detection Theory and Object Tracking," *International Journal on Applications in Information and Communication Engineering*, vol. 2, no. 4, pp. 57–61, 2016.
- [27] X.-M. Xu, et al., "Vehicle detection algorithm based on codebook and local binary patterns algorithms," *Journal of Central South University*, vol. 22, no. 2, pp. 593–600, 2015.
- [28] O. Barnic, M. Van Droogenbroeck, "ViBE: A powerful random technique to estimate the background in video sequences," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2009, pp. 945–948.
- [29] B. Olivier, M. Van Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2010.
- [30] R. B. Devi, Y. J. Chanu, K. M. Singh, "A Survey on Different Background Subtraction Method for Moving Object Detection," *International Journal For Research in Emerging Science and Technology*, vol. 3, no. 10, 2016.
- [31] Y. Xu, J. Dong, B. Zhang, D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAA Transactions on Intelligence Technology*, vol. 1, no. 1, pp. 43–60, 2016.
- [32] N. K. Kanhere, S. T. Birchfield, "Real-Time Incremental Segmentation and Tracking of Vehicles at Low Camera Angles Using Stable Features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 148–160, 2008.
- [33] I. Lipovac, et al., "Experimental evaluation of vehicle detection based on background modelling in daytime and night-time video," *Croatian Computer Vision Workshop, CCVW 2014*, 2014.
- [34] N. Buch, et al., "Detection and classification of vehicles for urban traffic scenes," *2008 5th International Conference on Visual Information Engineering*, 2008.
- [35] B. T. Morris, M. M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 425–437, 2008.
- [36] A. Manzanera, " - background subtraction and the zipf law," *Iberoamerican Congress on Pattern Recognition*, Springer, 2007, pp. 42–51.
- [37] S. Toral, M. Vargas, F. Barrero, M. Ortega, "Improved sigma-delta background estimation for vehicle detection," *Electronics Letters*, Springer, 2007, pp. 42–51.
- [38] J. M. Milla, S. L. Toral, M. Vargas, F. J. Barrero, "Dual-rate background subtraction approach for estimating traffic queue parameters in urban scenes," *IET Intelligent Transport Systems*, vol. 7, no. 1, pp. 122–130, 2013.

- [39] C. Stauffer, W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, IEEE, vol. 2, 1999, pp. 246–252.
- [40] A. Nurhadiyatna, et al., "Background Subtraction Using Gaussian Mixture Model Enhanced by Hole Filling Algorithm (GMMHF)," *2013 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 2013, pp. 4006–4011.
- [41] P. K. Bhaskar, S.-P. Yong, "Image processing based vehicle detection and tracking method," *2014 International Conference on Computer and Information Sciences (ICCOINS)*, IEEE, 2014, pp. 1–5.
- [42] Y. B. R. Indrabayu, et al., "Vehicle detection and tracking using gaussian mixture model and kalman filter," *Computational Intelligence and Cybernetics (CYBERNETICSCOM), International Conference, Indonesia*, IEEE, 2013, pp. 4006–4011.
- [43] Z. Huang, et al., "Vehicle ROI Extraction Based on Area Estimation Gaussian Mixture Model," *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, IEEE, 2017, pp. 1–7.
- [44] K. Karman, "Moving object segmentation based on adaptive reference images," *Proc. of EUSIPCO*, Barcelona, 1990.
- [45] X. Wang, J. Sun, H. Peng, "Foreground object detecting algorithm based on mixture of gaussian and kalman filter in video surveillance," *Journal of Computers*, vol. 8, no. 3, pp. 693–700, 2013.
- [46] N. M. Oliver, B. Rosario, A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000.
- [47] C. Y., Prabha DR IS, "Moving vehicle detection using Kalman method," *International Journal of Advanced and Innovative Research*, vol. 1, no. 2, 2012.
- [48] S. Srilekha, G. Swamy, A. A. Krishna, "A Novel Approach for Detection and Tracking of Vehicles Using Kalman Filter," *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, 2015, pp. 234–236.
- [49] A. Tah, S. Roy, P. Das, A. Mitra, "Moving object detection and segmentation using background subtraction by Kalman filter," *Indian Journal of Science and Technology*, IEEE, 2015, pp. 234–236.
- [50] Z. Hu, G. Ye, et al., "PKU@TRECVID2009: Single-Actor and Pair-Activity Event Detection in Surveillance Video," *TRECVID 2009 workshop participants notebook papers*, 2009.
- [51] Y. Tian, Y. Wang, Z. Hu, T. Huang, "Selective eigen background for background modeling and subtraction in crowded scenes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1849–1864, 2013.
- [52] S. S. Begum, K. L. Reddy, "FPGA Implementation of Selective Eigen Background Using Background Subtraction Algorithm in Crowded Scenes," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, 2014.
- [53] O. H. Ching, "Optical Flow-Based Vehicle Detection and Tracking," Bachelor Thesis, Information Systems Engineering, University Tunku Abdul Rahman, Perak, Malaysia, 2010.
- [54] Y. Liu, Y. Lu, Q. Shi, J. Ding, "Optical Flow Based Urban Road Vehicle Tracking," *2013 Ninth International Conference on Computational Intelligence and Security*, IEEE, 2013, pp. 391–395.
- [55] P. K. Bhaskar, S.-P. Yong, L. T. Jung, "Enhanced and effective parallel optical flow method for vehicle detection and tracking," *2015 International Symposium on Mathematical Sciences and Computing Research (iSMSC)*, IEEE, 2015, pp. 138–143.
- [56] S. Indu, M. Gupta, A. Bhattacharyya, "Vehicle Tracking and Speed Estimation using Optical Flow Method," *International Journal of Engineering Science and Technology*, vol. 3, no. 1, pp. 429–434, 2011.
- [57] D. G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, IEEE, 1999, pp. 1150–1157.
- [58] S. Agarwal, A. Awan, D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004.
- [59] X. Ma, W. E. L. Grimson, "Edge-based rich representation for vehicle classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE, 2005, pp. 1185–1192.
- [60] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, IEEE, vol. 1, 2001, pp. I–I.
- [61] S. ElKerdawy, A. Salaheldin, M. ElHelw, "Vision-based scale-adaptive vehicle detection and tracking for intelligent traffic monitoring," *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, IEEE, 2014, pp. 1044–1049.
- [62] N. Miller, et al., "A Hidden Markov Model for Vehicle Detection and Counting," *2015 12th Conference on Computer and Robot Vision*, IEEE, 2015, pp. 269–276.
- [63] H. Bai, J. Wu, C. Liu, "Motion and haar-like features based vehicle detection," *2006 12th International Multi-Media Modelling Conference*, IEEE, 2006.
- [64] B. F. Momin, T. M. Mujawar, "Vehicle detection and attribute based search of vehicles in video surveillance system," *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, IEEE, 2015, pp. 1–4.
- [65] T.-T. Nguyen, T. T. Nguyen, "A real time license plate detection system based on boosting learning algorithm," *2012 5th International Congress on Image and Signal Processing*, IEEE, 2012, pp. 819–823.
- [66] B. Zhang, Y. Zhou, "Reliable vehicle type classification by Classified Vector Quantization," *2012 5th International Congress on Image and Signal Processing*, IEEE, 2012, pp. 1148–1152.

- [67] X. Chen, Q. Meng, "Vehicle Detection from UAVs by Using SIFT with Implicit Shape Model," *2013 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 2013, pp. 3139–3144.
- [68] K. Mu, F. Hui, X. Zhao, "Multiple Vehicle Detection and Tracking in Highway Traffic Surveillance Video Based on SIFT Feature Matching," *Journal of Information Processing Systems*, vol. 12, no. 2, pp. 183–195, 2016.
- [69] J.-W. Hsieh, L.-C. Chen, D.-Y. Chen, "Symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 6–20, 2014.
- [70] T. D. Gamage, J. G. Samarawickrama, A. Pasqua, "GPU based non-overlapping multi-camera vehicle tracking," *7th International Conference on Information and Automation for Sustainability*, IEEE, 2014, pp. 1–6.
- [71] N. Buch, J. Orwell, S. A. Velastin, "3D Extended Histogram of Oriented Gradients (3DHOG) for Classification of Road Users in Urban Scenes," *Proceedings of the British Machine Conference*, 2009, pp. 1–11.
- [72] M. Cheon, W. Lee, C. Yoon, M. Park, "Vision-Based Vehicle Detection System With Consideration of the Detecting Location," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1243–1252, 2012.
- [73] H. T. Niknejad, et al., "On-Road Multivehicle Tracking Using Deformable Object Model and Particle Filter With Improved Likelihood Estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 748–758, 2012.
- [74] B.-F. Wu, et al., "A Relative-Discriminative-Histogram-of-Oriented-Gradients-Based Particle Filter Approach to Vehicle Occlusion Handling and Tracking," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4228–4237, 2013.
- [75] T. Yawen and G. Jinxu, et al., "Research on Vehicle Detection Technology Based on SIFT Feature," *2018 8th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, IEEE, 2018, pp. 274–278.
- [76] T. Gao, et al., "Moving Vehicle Tracking Based on SIFT Active Particle Choosing," *ICONIP 2008: Advances in Neuro-Information Processing*, Springer, 2008, pp. 695–702.
- [77] H. Bay, et al., "SURF: Speeded Up Robust Features," *ECCV 2006: Computer Vision – ECCV 2006*, Springer, 2006, pp. 404–417.
- [78] L.-C. Chen, et al., "Real-time vehicle color identification using symmetrical surfs and chromatic strength," *2015 IEEE international symposium on circuits and systems (ISCAS)*, IEEE, 2015, pp. 2804–2807.
- [79] B. Momin, S. Kumbhare, "Real-time vehicle color identification using symmetrical SURFs and chromatic strength," *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2015, pp. 1–4.
- [80] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, IEEE, 2005, pp. 886–893.
- [81] S. Tuermer, F. Kurz, P. Reinartz, U. Stilla, "Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 6, pp. 2327–2337, 2013.
- [82] H. Huijie, et al., "The Moving Vehicle Detection and Tracking System Based on Video Image," *2013 Third International Conference on Instrumentation, Measurement, Computer, Communication and Control*, IEEE, 2013, pp. 1277–1280.
- [83] S. M. Elkerdawi, R. Sayed, M. ElHelw, et al., "Real-Time Vehicle Detection and Tracking Using Haar-Like Features and Compressive Tracking," *ROBOT2013: First Iberian Robotics Conference*, Springer, 2014, pp. 381–390.

BIOGRAPHIES OF AUTHORS



Aymen Fadhil Abbas completed his bachelor degree in 2011 from University of Technology, Baghdad, Iraq. Received his M.Sc. Degree from Universiti Teknologi Malaysia, Johor Bahru, Malaysia in 2014. He is currently working toward the Ph.D. degree in computer and microelectronic systems with Universiti Teknologi Malaysia.

His research interests include intelligent surveillance systems, computer vision, intelligent transportation systems and vehicle make and model recognition.



Usman Ullah Sheikh received his Ph.D. degree (2009) in image processing and computer vision from Universiti Teknologi Malaysia. His research work is mainly on computer vision, machine learning, and embedded systems design.

He is currently a Senior Lecturer at Universiti Teknologi Malaysia.



Fahad Taha AL-Dhief received the B.S. in Software Engineering from Imam Jaafar Al-Sadiq University, Iraq in 2013 and also received M.S. in Computer Science from University Kebangsaan Malaysia, Malaysia in 2016. Currently, he is a Ph.D. student at School of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia. His research interests are sensor network, routing protocols, mobile ad-hoc network, and social network, Internet of things, machine learning, artificial neural networks, deep learning and location-based service.

He is an active student member of IEEE, and a member of IEEE Communications Society.



Mohd Norzali Haji Mohd He received B.Eng., M.Eng. from Fukui University, Japan in 2002 and 2004 respectively. In April 2015, he received Ph.D. from the Department of Information Sciences and Biomedical Engineering, Kagoshima University, Japan. His research interests are computer vision, robot vision, pattern recognition, and image processing.

Mohd Norzali is currently working as Senior Lecturer, Faculty Lab Manager (HoD) and Former Industrial Training Coordinator at the Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM).